

**Dependent Database Information
Is Complete and Examination Cases
Are Accurately Scored**

March 2003

Reference Number: 2003-40-091

This report has cleared the Treasury Inspector General for Tax Administration disclosure review process and information determined to be restricted from public release has been redacted from this document.



DEPARTMENT OF THE TREASURY
WASHINGTON, D.C. 20220

INSPECTOR GENERAL
for TAX
ADMINISTRATION

March 27, 2003

MEMORANDUM FOR COMMISSIONER, WAGE AND INVESTMENT DIVISION

Gordon C. Milbourn

FROM: Gordon C. Milbourn III
Acting Deputy Inspector General for Audit

SUBJECT: Final Audit Report - Dependent Database Information
Is Complete and Examination Cases Are Accurately Scored
(Audit # 200240075)

This report presents the results of our review to determine whether the Internal Revenue Service's (IRS) Dependent Database, used to identify Earned Income Tax Credit (EITC) overclaims, is complete and accurate and has adequate controls. This audit was conducted as a follow-up to a prior Treasury Inspector General for Tax Administration audit.¹

In summary, the IRS has instituted adequate controls to help ensure the completeness and accuracy of the Dependent Database. The database includes data from all 50 states and the District of Columbia. Criteria from the previous EITC examination selection methods have also been incorporated into the Dependent Database. In addition, analyses of a judgmental sample of database records revealed no instances of incomplete or inaccurate data in the database, and an analysis of a randomly selected sample of Dependent Database returns revealed no instances of inaccurate scores. However, validation of the accuracy and currency of data from other Federal and state agencies used in the Dependent Database was outside the scope of this review.

No recommendations were made in this report. However, we provided a copy of the report to IRS management for comments prior to issuance. IRS management decided

¹ *The Implementation of a New Audit Selection Program for Earned Income Credit Filers Could Result in Significant Taxpayer Burden* (Reference Number 2001-40-102, dated July 2001).

not to provide any written comments since the report was positive and had no recommendations requiring their action.

Copies of this report are also being sent to the IRS managers who are affected by the report finding. Please contact me at (202) 622-6510 if you have questions or Michael R. Phillips, Assistant Inspector General for Audit (Wage and Investment Income Programs), at (202) 927-0597.

**Dependent Database Information Is Complete
and Examination Cases Are Accurately Scored**

Table of Contents

Background	Page 1
The Dependent Database Has Adequate Controls to Ensure Completeness and Accuracy.....	Page 3
Appendix I – Detailed Objective, Scope, and Methodology	Page 7
Appendix II – Major Contributors to This Report.....	Page 9
Appendix III – Report Distribution List	Page 10
Appendix IV – The Dependent Database Scoring Program	Page 11

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

Background

The Congress has long been concerned with the administration of the Earned Income Tax Credit (EITC) program. Erroneous EITC claims are a source of significant loss of revenue for the Federal Government. The Internal Revenue Service (IRS) estimated that between \$8.5 and \$9.9 billion (27.0 percent to 31.7 percent) of the \$31 billion in EITC claimed for Tax Year 1999 were in error.¹ One of the main causes for these errors was taxpayers claiming children who did not meet the qualifications for the EITC. Children used by taxpayers to qualify for the EITC must meet relationship, age, and residency tests. In addition, the child used to qualify for the EITC cannot be the qualifying child of another person with a higher modified adjusted gross income,² and the taxpayers claiming the EITC cannot be used as qualifying children of other taxpayers claiming the EITC.³

In an attempt to address EITC errors, the Congress passed the Taxpayer Relief Act of 1997 (TRA 97),⁴ which provided a means for the IRS to improve its examination selection process. The TRA 97 includes a provision that gives the Department of the Treasury access to data collected by the Department of Health and Human Services (HHS). The HHS data include information about the person(s) with whom a child resides. Because residency is a requirement to claim the EITC, the HHS data would help the IRS identify a child's residence for determining entitlement to an EITC claim. As a result of the TRA 97, the IRS revised its examination selection program to incorporate the HHS Federal Case Registry of Child Support Orders (Federal Case Registry) into a computer system known as the Dependent Database. This Federal Case Registry contains

¹ The IRS' "Compliance Estimates for Earned Income Tax Credit on 1999 Returns," dated February 28, 2002. Details may not add due to rounding.

² For most people, modified adjusted gross income is the same as adjusted gross income but can include adding certain losses; tax exempt interest; and certain pension, annuity, or individual retirement arrangement distributions that are partly non-taxable.

³ Earned Income Credit [For use in preparing 2001 Returns] (Publication 596).

⁴ Taxpayer Relief Act of 1997, Pub. L. No. 105-34, 111 Stat. 961 §§ 1090(a) and 1090(b).

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

information pertaining to the residency and support of a child, including information on the custodial parent, non-custodial parent, and anyone who has been named as the father but for whom paternity has not been established.

The IRS uses the Dependent Database to identify and select for examination taxpayers with possible erroneous EITC claims. During initial processing, the Dependent Database Scoring Program analyzes tax returns that have claimed at least one EITC qualifying child or dependent child. Using data from several sources, it analyzes each tax return for criteria that indicate the taxpayer might not be eligible for the EITC and assigns a numeric value to each criterion. The Dependent Database then produces an overall score for the return. Based on resources available to conduct examinations, the IRS selects certain types and quantities of returns for pre-refund examinations to verify the taxpayers' eligibility for the EITC. Appendix IV provides additional information on the Dependent Database.

In Fiscal Year 2001, the Treasury Inspector General for Tax Administration conducted an audit of the Dependent Database and determined that the database did not include information from all 50 states and the District of Columbia.⁵ At the time of the audit, not all states were participating in the HHS Federal Case Registry. We recommended that the IRS coordinate with the HHS to obtain the Federal Case Registry data from all states. We also raised concerns about using only the Federal Case Registry data as examination selection criteria and recommended that the IRS incorporate rules from both its original selection program, known as the Electronic Fraud Detection System, and the Federal Case Registry into the Dependent Database. The Electronic Fraud Detection System uses common characteristics from past erroneous EITC claims as the basis for selecting taxpayers for examination.

We conducted this audit to determine if the IRS has made the suggested improvements to the Dependent Database and

⁵ *The Implementation of a New Audit Selection Program for Earned Income Credit Filers Could Result in Significant Taxpayer Burden* (Reference Number 2001-40-102, dated July 2001).

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

to assess the completeness and accuracy of the database. Complete and accurate, in this context, means that the data in the Dependent Database agreed with the data in the source files. Validation of the accuracy and currency of data from other Federal and state agencies used in the database was outside the scope of this review. We did not determine if the database is employing the most efficient examination selection criteria, nor did we determine if it uses all information available to the IRS when scoring and selecting returns for examination.

This audit was performed between August 2002 and January 2003. The review included visits to the Office of Compliance, Strategy and Selection, in the Wage and Investment Division Headquarters. We also visited the Corporate Data and Systems Management Division in the Headquarters Office of Modernization, Information Technology and Security Systems. The audit was conducted in accordance with *Government Auditing Standards*. Detailed information on our audit objective, scope, and methodology is presented in Appendix I. Major contributors to the report are listed in Appendix II.

The Dependent Database Has Adequate Controls to Ensure Completeness and Accuracy

The data contained in the Dependent Database are complete and accurate and have adequate controls. In addition, the Dependent Database is accurately applying the examination selection criteria and computing the related examination score.

The Dependent Database is complete

The Federal Case Registry data for all 50 states and the District of Columbia are included in the Dependent Database. In a prior report, we determined that Federal Case Registry data from nine states and the District of Columbia were not included in the database.

With all states included in the Dependent Database, the IRS was able to identify approximately 330,000 additional returns for pre-refund examinations between January and August 2002. The Federal Case Registry data, in accordance with the IRS requirement to provide the most current information for use in selecting returns for examination, are also updated periodically.

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

We raised concerns in the prior report that the IRS was using only the Federal Case Registry data as examination selection criteria and recommended the database incorporate criteria from the Electronic Fraud Detection System. The IRS has incorporated the Electronic Fraud Detection System criteria into the examination selection criteria for the Dependent Database. Returns are being selected for examination based on every set of criteria, including the Electronic Fraud Detection System criteria.

The Dependent Database contains information accurately transferred from source files, and the computer program accurately evaluates returns

A review of a judgmental sample of 70 records in the Dependent Database showed that source data are accurately transferred to the database. The contents of selected data fields in the records were compared to source data in the IRS' computer system, Social Security Administration data, and Federal Case Registry data. In all 70 records, data in the Dependent Database matched the applicable source data. A list of the source information used in the Dependent Database processing can be found in Appendix IV.

An additional evaluation of 20 of the 70 sample records showed that the Dependent Database is accurately applying the examination selection criteria and computing the related examination score. The examination selection criteria and scoring process outlined in the Dependent Database system documentation was manually applied to each of the 20 database records reviewed. In all 20 records, the score computed manually matched the score computed by the database.

While the Dependent Database is complete and accurate, there are several discrepancies in the current system documentation. Errors include typographical errors, incomplete or inaccurate data definitions, and inaccurate process descriptions. The IRS advised us that the discrepancies resulted because system documentation had not been updated when programming changes were made.

A review of the database computer program found that all programming changes indicated by the IRS have been made

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

in the computer program and that inaccurate or incomplete computer programming codes in the documentation were typographical errors or instances of incomplete updates to the Dependent Database documentation. The IRS advised us that the system documentation would be updated for the processing of Tax Year 2002 returns.

Controls help to ensure the completeness and accuracy of the Dependent Database

Prior to implementing the system nationwide, the IRS used the following testing techniques to ensure the Dependent Database was functioning as intended:

- Systems Acceptability Testing – Testing conducted using test data to ensure that the computer programs were operating as intended and were correctly evaluating and selecting returns.
- Pilot Testing – Testing conducted in a production environment on a limited basis.

During processing cycles, the IRS continues to test the Dependent Database using the following techniques:

- Real-time Testing – Testing conducted as returns are being processed to determine how well the evaluation criteria are identifying examination issues that result in an increase in tax liability.
- On-line Feedback – Obtaining ongoing feedback from field examination units during the processing cycle to surface any systemic problems that may be occurring.

Because the Federal Case Registry is received from sources outside the IRS, the IRS cannot attest to the accuracy of these data. However, to help ensure the Federal Case Registry's reliability, the IRS analyzes the data for reasonableness before including them in the Dependent Database. For example, data from the Federal Case Registry are analyzed to eliminate duplicate records. The Federal Case Registry data are also evaluated for inconsistencies in the data that would indicate the data are inaccurate, such as a child's birth date being prior to a parent's birth date.

**Dependent Database Information Is Complete
and Examination Cases Are Accurately Scored**

In addition, to safeguard against a taxpayer being selected for multiple examinations of the same tax return, the Dependent Database does not allow the same return to be considered for examination more than once. The system also does not allow any return previously examined to be entered into the Dependent Database.

Detailed Objective, Scope, and Methodology

The overall objective was to determine whether the Internal Revenue Service's (IRS) Dependent Database, used to identify Earned Income Tax Credit (EITC) overclaims, is complete and accurate and has adequate controls.

We conducted the following tests to accomplish the objective:

- I. To determine if the IRS has adequate controls in place to ensure the Dependent Database was complete, we:
 - A. Interviewed IRS personnel in the Wage and Investment (W&I) Division, Office of Compliance, Strategy and Selection; the EITC Project Office; and the Modernization, Information Technology and Security (MITS) Services organization and obtained pertinent documentation including the Dependent Database Functional Specification Package for Processing Year 2002.
 - B. Obtained access to the Dependent Database computer programs and an electronic copy of the database. We analyzed the Dependent Database data to ensure that the Federal Case Registry of Child Support Orders (Federal Case Registry)¹ data were present for all 50 states and the District of Columbia. We also verified that the database included returns scored with Electronic Fraud Detection System rules.² We did not validate the accuracy and currency of data from other Federal and state agencies used in the database because that was outside the scope of this review.
- II. To determine if the IRS has adequate controls in place to ensure the information contained in the Dependent Database is accurate, we:
 - A. Interviewed personnel in the W&I Division, Office of Compliance, Strategy and Selection, and the MITS Services organization and obtained pertinent documentation. We also determined how the IRS ensures the data contained in the Federal Case Registry are reliable. We did not validate the accuracy and currency of data from other Federal and state agencies used in the database because that was outside the scope of this review.
 - B. Evaluated a judgmental sample of 70 records from the Dependent Database to verify that source data are accurately transferred to the database. The Dependent Database

¹ A computer file obtained from the Department of Health and Human Services that provides information on the custody of a child.

² Examination selection rules used prior to the Dependent Database. These rules are based on characteristics commonly found on returns with erroneous EITC claims.

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

included 2,874,550 returns. Judgmental sampling was used to validate the data and establish an expected error rate for statistical sampling. The established error rate was zero; therefore, we did not expend additional audit resources to conduct a statistical sample. We conducted additional testing of 20 of the 70 records to verify that the Dependent Database computer program accurately scores returns for examination selection.

The 70 returns were selected for review using 2 methods. We selected the first 25 records in the Scored Table and the first 25 records in the Selected Table³ within the Dependent Database. We then generated 20 random numbers using a random number computer program. These random numbers were used to randomly select for review the corresponding numbered record in the Selected Table within the Dependent Database.

- C. Reviewed the Dependent Database system documentation and computer programs to determine whether programming changes were reflected in the computer code.

³ The Dependent Database is a relational database consisting of multiple sets of data or tables.

Major Contributors to This Report

Michael R. Phillips, Assistant Inspector General for Audit (Wage and Investment Income Programs)

Augusta R. Cook, Director

Deann L. Baiza, Audit Manager

Areta G. Heard, Senior Auditor

Doris J. Hynes, Senior Auditor

James M. Traynor, Senior Auditor

Robert J. Carpenter, Senior Information Technology Specialist

Report Distribution List

Acting Commissioner N:C
Deputy Commissioner, Modernization, Information Technology and Security Services M
Chief, Information Technology Services M:I
Director, Compliance W:CP
Director, Strategy and Finance W:S
Earned Income Tax Credit Program Manager W:EITC
Chief Counsel CC
National Taxpayer Advocate TA
Director, Legislative Affairs CL:LA
Director, Office of Program Evaluation and Risk Analysis N:ADC:R:O
Office of Management Controls N:CFO:AR:M
Audit Liaisons:
 Chief, Customer Liaison S:COM
 Program/Process Assistant Coordinator, Wage and Investment Division W:HR
 Director, Business Systems Development M:I:B
 Analyst, Program Oversight and Coordination Office M:R:PM:PO

The Dependent Database Scoring Program

The Dependent Database system was developed to add child custody and support data, acquired from the Department of Health and Human Services (HHS), to the processing of individual tax returns. The Internal Revenue Service (IRS) uses the Dependent Database scoring program to identify and select for examination taxpayers with possible erroneous Earned Income Tax Credit (EITC) claims. Between January and August 2002, the Dependent Database scoring program identified approximately 2.8 million returns for possible pre-refund examination. Of those, approximately 168,000 were selected for examination.

During initial tax return processing, all tax returns that have claimed at least one EITC qualifying child or dependent child are evaluated by this system. Using several data sources, the system analyzes the return for specific criteria. These criteria are based on characteristics that would indicate the taxpayer might not be eligible for the EITC.

There are 25 different sets of criteria used in the Dependent Database scoring program. A numeric value is assigned to each set of criteria, and each of the 25 criteria is applied to every tax return. The Dependent Database produces an overall score for the return based on the outcome of this analysis.

After all tax returns are scored, certain types and quantities of returns are selected for pre-refund examinations based on IRS resources available. If a return is selected for examination, the taxpayer's refund is "frozen" until the examination is completed and all questionable information is verified.

Internal and external data used in the Dependent Database scoring program include:

- Generalized Mainline Framework 15 – Primary file used in the Dependent Database that includes United States (U.S.) Individual Income Tax Return (Form 1040 series) information for each taxpayer.
- Federal Case Registry of Child Support Orders – Database created using information reported from state agencies and the District of Columbia to the HHS, including the identity of the custodial parent, the non-custodial parent, and any person who may be recognized as a parent, such as a legal guardian.
- Kidlink – Database containing information about a child's mother and father as reported to the Social Security Administration (SSA).
- DM1 – Database containing vital statistics for the entire U.S. population as reported to the SSA, such as the date of birth or date of death of an individual.

Dependent Database Information Is Complete and Examination Cases Are Accurately Scored

- DUPTIN – Database containing Social Security Numbers (SSN) and their usage in the IRS' tax return processing system.
- National Account Profile – Database containing taxpayer-identifying information including a taxpayer's SSN, name and address, and information pertaining to the taxpayer's spouse.
- Individual Returns Transaction File – File that contains tax return information that has completed the IRS' tax return processing system.
- Duplicate Direct Deposit – Database containing bank account information that associates all taxpayers who used the account for a direct deposit refund.
- Individual Master File On-Line – Files containing the on-line version of taxpayer information and tax return information.

While the primary focus of the Dependent Database process is to verify whether taxpayers that claim children to receive the EITC are meeting residency and/or relationship requirements, other issues are also examined. Those issues include:

- Duplicate claims of children (for the EITC or dependent purposes) by two or more taxpayers.
- Invalid filing status for the EITC.
- Deceased qualifying child or dependent.